Weekly Report

2018.0925-2018.0930

1. This Week

Deep Learning Power Grid Program

1.We finished 35 rounds of training (250 rounds in total). The accuracy on the training set has increased to over 96%, but the accuracy on the validation set maintains 15%. It looks like overfitting. We found that once we decrease the learning rate, the accuracy on the validation set will be significantly improved. So we transfer the data to the SSD and re-start the training, during which we manually adjust the learning rate.

2.We performed tensor decomposition (both cp decomposition and tucker decomposition) on a 36 node dataset. There exists the following problems:

- Features extracted by tensor decomposition are very similar in the same vector (rank = 1), making it hard to distinguish patterns in the dataset. When we try CP decomposition (rank = 3), the vector we get at the second time have great differences. Actually, the first vector should indicate the most representative pattern, but the second one presents patterns when the most representative pattern is removed from the dataset. It might be explained in this way (but I'm not sure about this)： the first vectors represent the stable patterns near the reference value, when the reference value is removed, the second vectors represents the patterns described by the difference, so it varies greatly.
- If we want to get a good initial value of tensor decomposition, we need to perform SVD to decompose feature matrices. But SVD cannot be performed on large matrix efficiently (it costs huge time and storage). So it might be difficult to apply tensor decomposition to large datasets. The 2,000 node dataset we got might be its limit, but we haven't tried.

3.I read the vector learning paper this week. And I'm thinking if we can encode a vector representation for each data sample (multi-variate time series sets). The length of the feature vectors and the curse of dimensionalities make it difficult to use other methods (except deep learning) to analyze simulation data. But if we can find a compact vector representation for each data sample, it might make things different.

Wavelines Revision

1.Revise the representation and visual analysis part of the paper:

- Move all visual designs to the representation section
- Shrink the task section
- Turn the original visual design section to explain the visual analysis pipeline

Working Hour: (except nap and eat time)

8-9 hours / day（Except Tuesday）

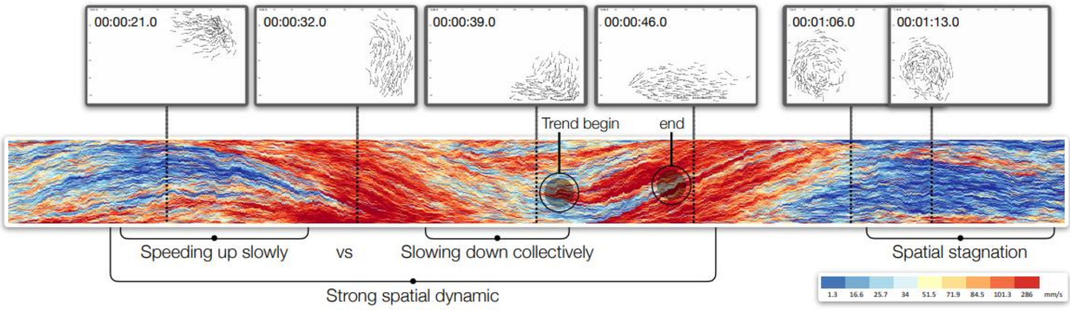
6 hours on Tuesday（Back to Hangzhou on Tuesday afternoon）

Total Working hour this week: 49 hours

Paper Reading

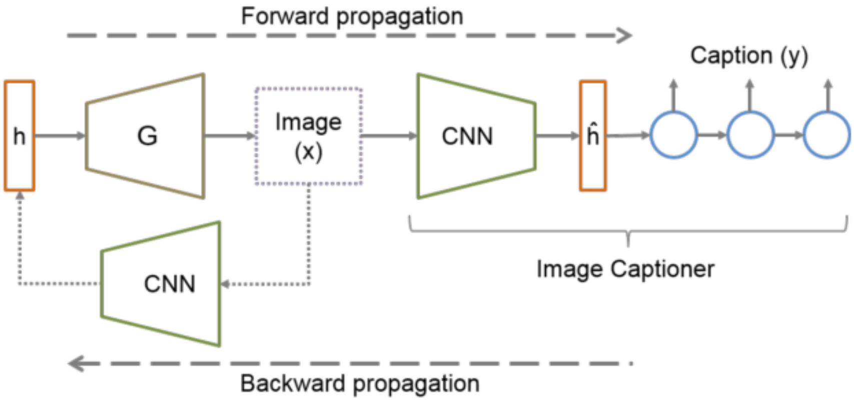
1. MotionRugs: Visualizing Collective Trends in Space and Time

This paper proposes a visualization technique to provide an overview of spatio-temporal data. It abstracts the spatial dimension to one dimensional slice by using space filling curves and spatial tree indices. The extracted on dimensional vectors are then horizontally aligned to produce an overview, with each pixel represent an entity at a time stamp. The space filling curves and spatial tree indices used in this paper are evaluated to examine whether the space segmentation method influences the visual representation.



2.Vector Learning for Cross Domain Representations

This paper proposes a model to embed text and image to the same vector space and realizes the task of generating an image according to a given text. It requires that the vector representation of the generated image and the given text to be similar. The proposed model comprises of three modules: a generator G, a CNN encoder and a image captioner model. The CNN encoder extracts the feature vector of a given image. The image captioner generates a text description of the image. After a backpropagation process, a reconstructed feature vector is used in the generator to generate a new image. This process is gradually updated.

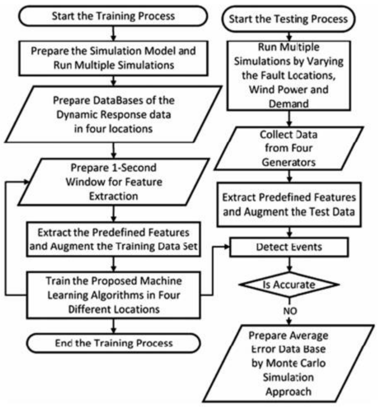


3. Dynamic Event Detection Using a Distributed Feature Selection Based Machine Learning Approach in a Self-Healing Microgrid (IEEE Power System 2018)

This paper is very similar to what we are doing now, the only difference is that it only uses the generator data.

It proposes an algorithm to detect dynamic events by applying a feature selection based machine learning approach. It is mostly based on one type of three-phase fault which leads to rotor angle instability. The proposed method detects features from the generator data

and then applies a multiclass classification algorithm to the feature data. Each class represents one dynamic event taking place. It starts with a data preparation stage where a Monte Carlo-based simulation is carried out with the four fault locations near to the wind power plant, solar plant, hydro and diesel power plants. Then it prepares a dynamic response database. The database contains the pre-fault, during fault and post-fault dynamic responses of each generator. Then it extracts predefined features and augment the dataset for training machine learning algorithms placed near each of the generators.



4. Real-Time Event Identification Through Low-Dimensional Subspace Characterization of High-Dimensional Synchrophasor Data

This paper proposes a real-time event identification method. The central idea is to characterize an event by the low-dimensional subspace spanned by the dominant singular vectors of a data matrix that contains the spatial–temporal block data. The subspace representation is robust to system initial conditions and characterizes the system dynamics. A dictionary of subspaces that corresponds to different events are established offline, and an event is identified online with the most similar event in the dictionary through subspace comparison.

5. Wide-Area Monitoring of Power Systems Using Principal Component Analysis and k-Nearest Neighbor Analysis

This paper combines PCA with kNN to detect and locate power system disturbances in real time. It builds two new system-wide monitoring statistics, which not only monitors a large number of variables but also reduces the masking effect of the oscillatory trends and noise. A contribution plot strategy quantifying the contributions of variables to the anomaly of the new statistics is developed. It identifies the variables affected most by the detected disturbance and thus can provide a significant reference for finally locating the detected disturbance.

2. Progress

Task	Progress	Time
Wavelines Revision	1.Finish the second version of	ASAP

	revision.	
Power Grid Paper with Deep Learning	1.Train the 2,000 node data (100,000 samples) 2.Perform tensor decomposition	12.15
SQC paper	Delayed	-